MODOMICS: a database of RNA modification pathways—2013 update

Magdalena A. Machnicka[1], Kaja Milanowska[1,2], Okan Osman Oglou[3], Elzbieta Purta[1], Malgorzata Kurkowska[1], Anna Olchowik[1], Witold Januszewski[1], Sebastian Kalinowski[2], Stanislaw Dunin-Horkawicz[1], Kristian M. Rother[1,2], Mark Helm[3], Janusz M. Bujnicki[1,2,*] and Henri Grosjean[4,*]

[1]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. Ks. Trojdena 4, PL-02-109 Warsaw, [2]Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 89, PL-61-614 Poznan, Poland, [3]Institut für Pharmazie und Biochemie, Johannes Gutenberg-Universität, Staudinger Weg 5, D-55128 Mainz, Germany and [4]Centre de Génétique Moléculaire, UPR 3404, CNRS, Université Paris-Sud, FRC 3115, 1 avenue de la Terrasse, 91190 Gif-sur-Yvette, France

ABSTRACT

MODOMICS is a database of RNA modifications that provides comprehensive information concerning the chemical structures of modified ribonucleosides, their biosynthetic pathways, RNA-modifying enzymes and location of modified residues in RNA sequences. In the current database version, accessible at http://modomics.genesilico.pl, we included new features: a census of human and yeast snoRNAs involved in RNA-guided RNA modification, a new section covering the 5′-end capping process, and a catalogue of 'building blocks' for chemical synthesis of a large variety of modified nucleosides. The MODOMICS collections of RNA modifications, RNA-modifying enzymes and modified RNAs have been also updated. A number of newly identified modified ribonucleosides and more than one hundred functionally and structurally characterized proteins from various organisms have been added. In the RNA sequences section, snRNAs and snoRNAs with experimentally mapped modified nucleosides have been added and the current collection of rRNA and tRNA sequences has been substantially enlarged. To facilitate literature searches, each record in MODOMICS has been cross-referenced to other databases and to selected key publications. New options for database searching and querying have been implemented, including a BLAST search of protein sequences and a PARALIGN search of the collected nucleic acid sequences.

INTRODUCTION

During the course of RNA maturation, various enzymes are able to introduce chemical modifications into ribonucleotide residues. Chemical alteration may occur in the base, at the 2′-hydroxyl of the ribose, or both. Indeed, many modified residues in fact correspond to intermediates in the sequential, multistep formation of hypermodified nucleotides (1). There are also modifications whose biosynthesis starts outside the RNA. For example, queuosine derivatives arise from azaguanine, in which extra chemical groups are attached to C7 (rather than N7 as in guanine). These compounds are introduced into RNA by transglycosylation involving the replacement of an original unmodified guanine by a modified pre-Q base (2). Likewise, in some positive-sense RNA viruses, such as those of the family *Alphaviridae*, the mRNA cap is formed by nucleotidyltransfer involving a GTP nucleotide pre-modified to N^7-methyl-GTP (m^{7}GTP) (3). The variety of chemical groups and their positions in natively modified nucleosides in RNA is illustrated in Figure 1.
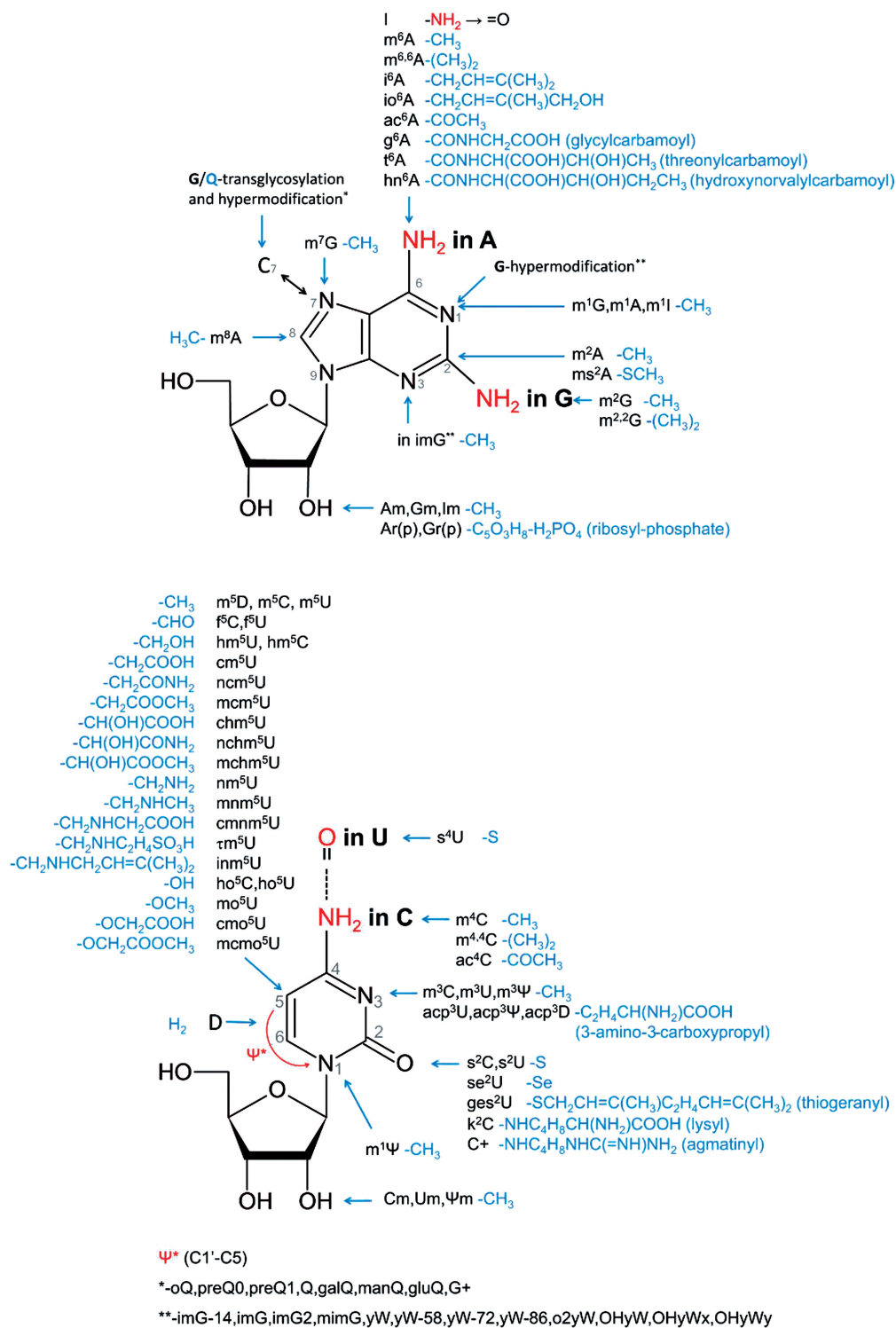
The location, abundance and distribution of various types of modification vary greatly between different RNA molecules, organisms and organelles. Physiological environment and growth conditions of the cell also affect the pattern of RNA modification and/or the degree of individual modifications among different RNA molecules of the same type (4). While the majority of modified nucleosides are present in transfer and ribosomal RNAs of

I $-NH_2 \rightarrow =O$
m^6A $-CH_3$
$m^{6,6}A$ $-(CH_3)_2$
i^6A $-CH_2CH=C(CH_3)_2$
io^6A $-CH_2CH=C(CH_3)CH_2OH$
ac^6A $-COCH_3$
g^6A $-CONHCH_2COOH$ (glycylcarbamoyl)
t^6A $-CONHCH(COOH)CH(OH)CH_3$ (threonylcarbamoyl)
hn^6A $-CONHCH(COOH)CH(OH)CH_2CH_3$ (hydroxynorvalylcarbamoyl)

G/**Q**-transglycosylation and hypermodification*

m^7G $-CH_3$ NH_2 **in A**

G-hypermodification**

m^1G, m^1A, m^1I $-CH_3$

$H_3C- m^8A$

m^2A $-CH_3$
ms^2A $-SCH_3$

NH_2 **in G** m^2G $-CH_3$
$m^{2,2}G$ $-(CH_3)_2$

in imG** $-CH_3$

HO

Am, Gm, Im $-CH_3$
$Ar(p), Gr(p)$ $-C_5O_3H_8-H_2PO_4$ (ribosyl-phosphate)

$-CH_3$ m^5D, m^5C, m^5U
$-CHO$ f^5C, f^5U
$-CH_2OH$ hm^5U, hm^5C
$-CH_2COOH$ cm^5U
$-CH_2CONH_2$ ncm^5U
$-CH_2COOCH_3$ mcm^5U
$-CH(OH)COOH$ chm^5U
$-CH(OH)CONH_2$ $nchm^5U$
$-CH(OH)COOCH_3$ $mchm^5U$
$-CH_2NH_2$ nm^5U
$-CH_2NHCH_3$ mnm^5U
$-CH_2NHCH_2COOH$ $cmnm^5U$
$-CH_2NHC_2H_4SO_3H$ τm^5U
$-CH_2NHCH_2CH=C(CH_3)_2$ inm^5U
$-OH$ ho^5C, ho^5U
$-OCH_3$ mo^5U
$-OCH_2COOH$ cmo^5U
$-OCH_2COOCH_3$ $mcmo^5U$

O **in U** $\leftarrow$ s^4U $-S$

NH_2 **in C** m^4C $-CH_3$
$m^{4,4}C$ $-(CH_3)_2$
ac^4C $-COCH_3$

H_2 D

Ψ^*

$m^3C, m^3U, m^3\Psi$ $-CH_3$
$acp^3U, acp^3\Psi, acp^3D$ $-C_2H_4CH(NH_2)COOH$ (3-amino-3-carboxypropyl)

s^2C, s^2U $-S$
se^2U $-Se$
ges^2U $-SCH_2CH=C(CH_3)C_2H_4CH=C(CH_3)_2$ (thiogeranyl)
k^2C $-NHC_4H_8CH(NH_2)COOH$ (lysyl)
$C+$ $-NHC_4H_8NHC(=NH)NH_2$ (agmatinyl)

HO

$m^1\Psi$ $-CH_3$

$Cm, Um, \Psi m$ $-CH_3$

Ψ^* (C1'-C5)

*$-oQ, preQ0, preQ1, Q, galQ, manQ, gluQ, G+$

**$-imG-14, imG, imG2, mimG, yW, yW-58, yW-72, yW-86, o2yW, OHyW, OHyWx, OHyWy$

Figure 1. The diversity of chemical groups introduced enzymatically during the process of RNA maturation at various locations of the four common ribonucleosides A, G, C and U linked via 5′- to 3′-phosphodiester bonds. Only conventional acronyms/symbols are shown. The complete scientific names of all modified nucleosides presented in this figure are provided in the MODOMICS database. Nucleosides in RNA can also be doubly or triply modified by the introduction of several of these chemical groups.

all types of cells, they have also been found to occur in small non-coding RNAs, such as spliceosomal (sn)RNAs, small nucleolar (sno)RNAs and more recently in regulatory RNAs, such as siRNAs, miRNAs and piRNAs (5,6).

The presence of modified bases in mRNAs and viral RNAs and their potential role in the regulation of gene expression has been recently intensively studied (7–9). New types of modifications have been found (10–16),

and biochemical and physiological roles have been revealed for many known modified ribonucleosides. Examples include the immune response to rRNA and tRNA methylation (17,18), the linkage between tRNA modification and host resistance to viral infection (19) and stress-induced cleavage of small RNAs (20). Many of these advances were driven by the use of synthetic RNA containing naturally occurring modified nucleotides. Moreover, numerous new RNA-modifying enzymes have been identified and characterized (21–26). To adequately represent this rapid accumulation of knowledge, we have added both to the variety and volume of data in the MODOMICS database. The most significant additions are: (i) snoRNAs linked to the corresponding modification sites in human and yeast RNAs; (ii) modifications in snRNAs and snoRNAs; (iii) update of the recently identified enzymes and pathways; (iv) a catalogue of 'building blocks' for the chemical synthesis of naturally occurring modified nucleosides. The implementation of BLAST (27) and PARALIGN (28) sequence search engines facilitates access to MODOMICS data on the level of protein and nucleic acid sequences. Finally, greater functionality has been added to the user interface.

DATABASE CONTENT

The MODOMICS database (http://modomics.genesilico .pl) has been developed to house and distribute collections of RNA modification pathways, chemical structures of modified nucleosides, sequences of modified RNAs, enzymes responsible for individual reactions and a catalogue of 'building blocks' for chemical synthesis of modified RNA. MODOMICS was created as a single resource to organize and present all these data in a convenient and straightforward way. Information about modified residues is also available in the RNMDB database (29). General-purpose pathway databases, such as REACTOME (30) also present some aspects of RNA modification pathways. However, MODOMICS is currently the most comprehensive source of information among all existing RNA modification databases.

MODIFICATIONS

At present, MODOMICS contains 144 different modifications that have been identified in RNA molecules [34 were added since the previous database release (31)]. A typical entry for a modified ribonucleoside contains information about its basic chemical properties, localization in known RNA molecule types, the phylogenetic distribution with respect to Domains of Life and known enzymes responsible for its biosynthesis. The list of modified nucleosides can be browsed by the modification names, the standard bases (A, G, C and U) from which they originate and the chemical groups they contain. The available details contain full and short names, the sum formula, PubChem ID and—to facilitate MS analyses of modified RNAs—the monoisotopic, HRMS and average masses. The chemical structures of the modified nucleosides are represented by 1D SMILE codes, 2D structure plots and

3D structures in the mol format displayed interactively on the website by a Jmol applet. Reactions linking a modified nucleoside to its precursor(s) and to hypermodifications are listed. Many of the products of modification reactions are substrates of further reactions, and the formation of hypermodified residues occurs in complex pathways, which are displayed as graphs. All modified nucleosides found in RNA structures deposited in the RCSB Protein Data Bank are also indicated and appropriate hyperlinks are provided.

PATHWAYS

MODOMICS comprises a collection of RNA modification pathways divided into six different categories according to their starting point: four categories correspond to the standard bases (A, G, C and U), another presents the incorporation and hypermodification pathway of queuosine and the other the modifications of the RNA 5′-cap. The pathway display in MODOMICS has undergone a radical overhaul. The new display, which is based on the Cytoscape Web network visualization tool (32), allows users to zoom and change the graph layout, and to save the obtained result as an image (png or svg), pdf or xml file. Pathway graphs are now easier to navigate and present information about the structures of modified nucleosides and the type of chemical reactions involved. Information about enzymatic transformations that have been verified experimentally (plain arrows) are distinguished from 'putative' reactions that are predicted but not yet experimentally confirmed (dashed arrows). Users can access detailed information about each reaction (including the type of transformation and enzymes responsible for its execution) by simply clicking the chosen arrow of the graph. The display of pathway graphs is currently supported for Google Chrome, Mozilla Firefox and for Microsoft Internet Explorer 9 (with Document Mode set to Internet Explorer 9 standards).

RNA SEQUENCES

MODOMICS provides a collection of modified RNA sequences of different types, such as tRNAs, rRNAs, snRNAs and snoRNAs. For families of homologous RNAs, multiple sequence alignments are available. Sequences are visualized with all modifications highlighted and linked to the corresponding modification record. The alignments can be displayed directly on the webpage or in a Jalview applet, or downloaded in plain text format. In comparison to the previous release of MODOMICS, the set of tRNA sequences has been updated and greatly expanded. In particular, in addition to the previously collected manually curated sequences, MODOMICS contains now all tRNAs imported from the Transfer RNA database (33), which were found to possess modified nucleosides (over 500 tRNA sequences in total). The secondary structure of tRNAs is also indicated now. rRNA sequence alignments and secondary structure indications have been updated as well, based on the data from the Comparative RNA Website (CRW) (34). MODOMICS

includes an arbitrarily selected subset of rRNA sequences representing different phylogenetic taxa, based on those from the CRW database. For these sequences, both the positions of modified residues and the identities of rRNA-modifying enzymes are known. The current release (as of September 2012) contains 10 SSU rRNA sequences (5 from Bacteria, 2 from Archaea and 3 from Eukaryota) and 9 LSU rRNA sequences (4 from Bacteria, 2 from Archaea and 3 from Eukaryota). For the cytoplasmic LSU in Eukaryota, 5.8S and 28S rRNAs are presented as a single-fused molecule, as in the CRW database. Currently, only one representative rRNA SSU and one LSU sequence per species is included. In future, we intend to expand this data set to include rRNAs from additional species, as well as to cover all rRNA variants encoded by a given genome. To map modifications onto rRNA sequences, we used data from the 'The Small Subunit rRNA Modification Database' (35), the 3D rRNA modification maps database (36) and the recently published data concerning modifications in both ribosomal subunits. Finally, we have improved MODOMICS by adding modified snRNA and snoRNA sequences. Unmodified snRNA and snoRNA sequences and alignments were obtained from the Rfam database (37). Positions of modifications in snRNA and snoRNA sequences were included based on the published data.

A new utility that allows mapping the modified positions on secondary structure diagrams of RNA molecules has been implemented. All modified positions from sequences collected in MODOMICS can be mapped onto reference diagrams based on the sequence alignments. For rRNAs, we used the structure of *Escherichia coli* SSU and LSU rRNAs obtained from the CRW as a reference. For tRNAs, we generated a consensus secondary structure diagram using VARNA (38), based on the data obtained from the Transfer RNA database. Graphics are presented using the JavaScript InfoVis Toolkit library (http://thejit.org/: Nicolas Garcia Belmonte). It is possible to map information from a user-selected set of sequences onto the diagram. In such a case, the percentage of modified ribonucleosides of any type in each alignment position is calculated and displayed. The resulting diagrams can be downloaded as image files.

PROTEINS

The MODOMICS database currently contains information about 274 proteins involved in RNA modification, both functional enzymes and protein co-factors necessary for multi-protein enzymatic activities. More than one hundred functionally and structurally characterized proteins have been added since the previous release (31) and the collection of protein sequences has been updated accordingly. We expanded the collection to include not only the functionally characterized RNA-modifying enzymes from *E. coli* and *Saccharomyces cerevisiae*, but also from other organisms, in particular if their crystal structures were available. 'Predicted' enzymes, whose activity has not been experimentally validated by genetic or biochemical methods, are currently excluded from MODOMICS. Enzymes that have been characterized biochemically *in vitro* or *in vivo*, but for which the corresponding genes have not been identified, are also excluded (although the corresponding reactions are collated).

The MODOMICS catalogue of proteins can be browsed by the source organism and/or type of the enzyme activity (methyltransferase, pseudouridine synthase, etc.). A list of matches can be further edited by the user, based on the features, such as name, position of modification, GI, COGs, PDB ID of structure, etc. At the level of individual protein entries, the database provides information about protein name(s), synonyms, amino acid sequence, corresponding ORF, modified RNA(s) and the position of the residues modified (if available). For proteins that are parts of enzymatic complexes, the name of the complex is provided. Reactions known to be catalysed by each protein with enzymatic activity are listed. Accession numbers from the Swiss-Prot (39) and GenPept (40) databases are provided and proteins with experimentally determined structures are linked to appropriate entries in the Protein Data Bank (41).

snoRNAs

As a new feature, we have included a census of human and yeast snoRNAs, involved in RNA-guided RNA modification by the C/D box and H/ACA box ribonucleoproteins, and we linked these snoRNAs to the corresponding modification sites in human and yeast RNAs. For this section of MODOMICS, we used data from the Saccharomyces Genome Database (42), the Yeast snoRNA Database (43) and snoRNA-LBME-db (44). The list of snoRNAs can be browsed by organism and/or type of modification found in the target position. Links to the HGNC database (45) and the Yeast snoRNA Database are provided for human and yeast snoRNAs, respectively.

BUILDING BLOCKS

In this MODOMICS database release, we have added a catalogue of 'building blocks' for the chemical synthesis of naturally occurring modified nucleosides. Each modification is uniquely characterized by its IUPAC name and CAS number, but more than one building block may be available for a given modification. The compilation is intended to facilitate solid phase synthesis of modified RNA, and thus to foster biophysical and biochemical studies. It provides a rapid overview of which modifications may be chemically incorporated with little or moderate synthetic effort, and inversely, which modifications remain attractive targets for synthetic bioorganic chemistry endeavours. The listing was compiled from the CAS database. As not all CAS entries are contained in PubMed, the relevant literature pertaining to synthesis and incorporation of building blocks is given. Its contents reflect the overall dominance of 'classical' phosphoramidite chemistry, featuring an acid labile 5′O-protecting group and a fluoride labile 2′O-protecting group. Protective groups that have proven useful in published syntheses are given, including conditions for chemical

deprotection after RNA synthesis. The protective groups themselves have separate entries, and further information is available on the reagent used for their introduction, and cross references to other building blocks containing the same protecting group.

SEARCH

New options for database searching and querying have been implemented, including a BLAST (27) search of protein sequences and a PARALIGN (28) search of nucleic acid sequences collected in MODOMICS, as well as a utility that sends a protein sequence from a MODOMICS entry to BLAST on the NCBI webserver. Hits and query-hit alignments resulting from MODOMICS searches can be downloaded in fasta format.

FUTURE PROSPECTS

The total number of confirmed modifications and RNA-modifying enzymes is growing continuously. Though there is a considerable amount of experimentally derived information available, there are still many modified positions in well-characterized RNA molecules for which the responsible enzymes are not known. New modified nucleosides are also being discovered, especially in RNA originating from more recently adopted model systems, such as extremophilic prokaryotes. Projects aimed at systematically studying genomes of related organisms holds promise of many more surprises [e.g. the recent initiative to sequence 5000 insects genomes (46)]. Thus, characterization of RNA modification pathways appears to be a moving target, and we appreciate feedback from users who bring newly discovered modifications and enzymes to our attention, and make suggestions of new methods of data presentation. In the future, we plan to further develop the graphical presentation of RNA modification data to allow comparative analysis of modification pathways and modification positions in sequences with respect to chosen taxa. An ultimate goal is to integrate MODOMICS with databases on other aspects of RNA metabolism. We also intend to link the information available in MODOMICS with 'RNAcentral', the planned database of RNA sequences (47).

AVAILABILITY

The data are freely accessible for research purposes at http://modomics.genesilico.pl. Most of the data are available for download in plain text formats. Modified nucleosides and building blocks are also available as structure files and images. Images of pathways are available for download from the web page in several formats. The pathway graphs can be also downloaded as an xml file (graphml format). Program code for parsing the plain text formats is available on request.

ACKNOWLEDGEMENTS

We thank Dr Murray Coles for critical reading of the article and Juliusz Stasiewicz for help in the curation of Modifications section of the database. We would like to thank all previous contributors to the MODOMICS database for their work that provided a solid basis for the present updates. We are indebted to the authors of the primary databases and services, whose content could be reused or linked to by MODOMICS. Last, but not least, we thank all the users of MODOMICS who provided feedback and made suggestions, and who cited MODOMICS in their publications.

REFERENCES

1. Grosjean,H. (2009) Nucleic acids are not boring long polymers of only four types of nucleotides: a guided tour. In: Grosjean,H. (ed.), *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin, pp. 1–18.
2. Iwata-Reuyl,D. (2008) An embarrassment of riches: the enzymology of RNA modification. *Curr. Opin. Chem. Biol.*, **12**, 126–133.
3. Ahola,T. and Kaariainen,L. (1995) Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl Acad. Sci. USA*, **92**, 507–511.
4. Globisch,D., Pearson,D., Hienzsch,A., Bruckl,T., Wagner,M., Thoma,I., Thumbs,P., Reiter,V., Kneuttinger,A.C., Muller,M. *et al.* (2011) Systems-based analysis of modified tRNA bases. *Angew Chem. Int. Ed. Engl.*, **50**, 9739–9742.
5. Montgomery,T.A., Rim,Y.S., Zhang,C., Dowen,R.H., Phillips,C.M., Fischer,S.E. and Ruvkun,G. (2012) PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1. *PLoS Genet.*, **8**, e1002616.
6. Kamminga,L.M., Luteijn,M.J., den Broeder,M.J., Redl,S., Kaaij,L.J., Roovers,E.F., Ladurner,P., Berezikov,E. and Ketting,R.F. (2010) Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J.*, **29**, 3688–3700.
7. Bodi,Z., Button,J.D., Grierson,D. and Fray,R.G. (2010) Yeast targets for mRNA methylation. *Nucleic Acids Res.*, **38**, 5327–5335.
8. Dominissini,D., Moshitch-Moshkovitz,S., Schwartz,S., Salmon-Divon,M., Ungar,L., Osenberg,S., Cesarkas,K., Jacob-Hirsch,J., Amariglio,N., Kupiec,M. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201–206.
9. Meyer,K.D., Saletore,Y., Zumbo,P., Elemento,O., Mason,C.E. and Jaffrey,S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell*, **149**, 1635–1646.
10. Havelund,J.F., Giessing,A.M., Hansen,T., Rasmussen,A., Scott,L.G. and Kirpekar,F. (2011) Identification of 5-hydroxycytidine at position 2501 concludes characterization of modified nucleotides in E. coli 23S rRNA. *J. Mol. Biol.*, **411**, 529–536.
11. Krog,J.S., Espanol,Y., Giessing,A.M., Dziergowska,A., Malkiewicz,A., de Pouplana,L.R. and Kirpekar,F. (2011) 3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine is one of two novel post-transcriptional modifications in tRNALys(UUU) from Trypanosoma brucei. *FEBS J.*, **278**, 4782–4796.

12. Chen,Y.G., Kowtoniuk,W.E., Agarwal,I., Shen,Y. and Liu,D.R. (2009) LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat. Chem. Biol.*, **5**, 879–881.

13. Kowtoniuk,W.E., Shen,Y., Heemstra,J.M., Agarwal,I. and Liu,D.R. (2009) A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc. Natl Acad. Sci. USA*, **106**, 7768–7773.

14. Mandal,D., Kohrer,C., Su,D., Russell,S.P., Krivos,K., Castleberry,C.M., Blum,P., Limbach,P.A., Soll,D. and RajBhandary,U.L. (2010) Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *Proc. Natl Acad. Sci. USA*, **107**, 2872–2877.

15. Ikeuchi,Y., Kimura,S., Numata,T., Nakamura,D., Yokogawa,T., Ogata,T., Wada,T., Suzuki,T. and Suzuki,T. (2010) Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. *Nat. Chem. Biol.*, **6**, 277–282.

16. Giessing,A.M., Jensen,S.S., Rasmussen,A., Hansen,L.H., Gondela,A., Long,K., Vester,B. and Kirpekar,F. (2009) Identification of 8-methyladenosine as the modification catalyzed by the radical SAM methyltransferase Cfr that confers antibiotic resistance in bacteria. *RNA*, **15**, 327–336.

17. Oldenburg,M., Kruger,A., Ferstl,R., Kaufmann,A., Nees,G., Sigmund,A., Bathke,B., Lauterbach,H., Suter,M., Dreher,S. *et al.* (2012) TLR13 recognizes bacterial 23S rRNA devoid of erythromycin resistance-forming modification. *Science*, **337**, 1111–1115.

18. Gehrig,S., Eberle,M.E., Botschen,F., Rimbach,K., Eberle,F., Eigenbrod,T., Kaiser,S., Holmes,W.M., Erdmann,V.A., Sprinzl,M. *et al.* (2012) Identification of modifications in microbial, native tRNA that suppress immunostimulatory activity. *J. Exp. Med.*, **209**, 225–233.

19. Maynard,N.D., Macklin,D.N., Kirkegaard,K. and Covert,M.W. (2012) Competing pathways control host resistance to virus via tRNA modification and programmed ribosomal frameshifting. *Mol. Syst. Biol.*, **8**, 567.

20. Schaefer,M., Pollex,T., Hanna,K., Tuorto,F., Meusburger,M., Helm,M. and Lyko,F. (2010) RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.*, **24**, 1590–1595.

21. Purta,E., O'Connor,M., Bujnicki,J.M. and Douthwaite,S. (2009) YgdE is the 2′-O-ribose methyltransferase RlmM specific for nucleotide C2498 in bacterial 23S rRNA. *Mol. Microbiol.*, **72**, 1147–1158.

22. Chatterjee,K., Blaby,I.K., Thiaville,P.C., Majumder,M., Grosjean,H., Yuan,Y.A., Gupta,R. and de Crecy-Lagard,V. (2012) The archaeal COG1901/DUF358 SPOUT-methyltransferase members, together with pseudouridine synthase Pus10, catalyze the formation of 1-methylpseudouridine at position 54 of tRNA. *RNA*, **18**, 421–433.

23. Wurm,J.P., Meyer,B., Bahr,U., Held,M., Frolow,O., Kotter,P., Engels,J.W., Heckel,A., Karas,M., Entian,K.D. *et al.* (2010) The ribosome assembly factor Nep1 responsible for Bowen-Conradi syndrome is a pseudouridine-N1-specific methyltransferase. *Nucleic Acids Res.*, **38**, 2387–2398.

24. Desmolaize,B., Fabret,C., Bregeon,D., Rose,S., Grosjean,H. and Douthwaite,S. (2011) A single methyltransferase YefA (RlmCD) catalyses both m5U747 and m5U1939 modifications in Bacillus subtilis 23S rRNA. *Nucleic Acids Res.*, **39**, 9368–9375.

25. Kempenaers,M., Roovers,M., Oudjama,Y., Tkaczuk,K.L., Bujnicki,J.M. and Droogmans,L. (2010) New archaeal methyltransferases forming 1-methyladenosine or 1-methyladenosine and 1-methylguanosine at position 9 of tRNA. *Nucleic Acids Res.*, **19**, 6533–6543.

26. Guy,M.P., Podyma,B.M., Preston,M.A., Shaheen,H.H., Krivos,K.L., Limbach,P.A., Hopper,A.K. and Phizicky,E.M. (2012) Yeast Trm7 interacts with distinct proteins for critical modifications of the tRNAPhe anticodon loop. *RNA*, **18**, 1921–1933.

27. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

28. Saebo,P.E., Andersen,S.M., Myrseth,J., Laerdahl,J.K. and Rognes,T. (2005) PARALIGN: rapid and sensitive sequence similarity searches powered by parallel computing technology. *Nucleic Acids Res.*, **33**, W535–W539.

29. Cantara,W.A., Crain,P.F., Rozenski,J., McCloskey,J.A., Harris,K.A., Zhang,X., Vendeix,F.A., Fabris,D. and Agris,P.F. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.*, **39**, D195–D201.

30. Croft,D., O'Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.

31. Czerwoniec,A., Dunin-Horkawicz,S., Purta,E., Kaminska,K.H., Kasprzak,J.M., Bujnicki,J.M., Grosjean,H. and Rother,K. (2009) MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res.*, **37**, D118–D121.

32. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.

33. Juhling,F., Morl,M., Hartmann,R.K., Sprinzl,M., Stadler,P.F. and Putz,J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.*, **37**, D159–D162.

34. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Muller,K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2.

35. McCloskey,J.A. and Rozenski,J. (2005) The Small Subunit rRNA Modification Database. *Nucleic Acids Res.*, **33**, D135–D138.

36. Piekna-Przybylska,D., Decatur,W.A. and Fournier,M.J. (2008) The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.*, **36**, D178–D183.

37. Gardner,P.P., Daub,J., Tate,J., Moore,B.L., Osuch,I.H., Griffiths-Jones,S., Finn,R.D., Nawrocki,E.P., Kolbe,D.L., Eddy,S.R. *et al.* (2011) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res.*, **39**, D141–D145.

38. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.

39. Magrane,M. and Consortium,U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database*, **2011**, bar009.

40. Sayers,E.W., Barrett,T., Benson,D.A., Bolton,E., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Federhen,S. *et al.* (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **39**, D38–D51.

41. Deshpande,N., Addess,K.J., Bluhm,W.F., Merino-Ott,J.C., Townsend-Merino,W., Zhang,Q., Knezevich,C., Xie,L., Chen,L., Feng,Z. *et al.* (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, **33**, D233–D237.

42. Cherry,J.M., Hong,E.L., Amundsen,C., Balakrishnan,R., Binkley,G., Chan,E.T., Christie,K.R., Costanzo,M.C., Dwight,S.S., Engel,S.R. *et al.* (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.

43. Piekna-Przybylska,D., Decatur,W.A. and Fournier,M.J. (2007) New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA*, **13**, 305–312.

44. Lestrade,L. and Weber,M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.

45. Bruford,E.A., Lush,M.J., Wright,M.W., Sneddon,T.P., Povey,S. and Birney,E. (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.*, **36**, D445–D448.

46. Robinson,G.E., Hackett,K.J., Purcell-Miramontes,M., Brown,S.J., Evans,J.D., Goldsmith,M.R., Lawson,D., Okamuro,J., Robertson,H.M. and Schneider,D.J. (2011) Creating a buzz about insect genomes. *Science*, **331**, 1386.

47. Bateman,A., Agrawal,S., Birney,E., Bruford,E.A., Bujnicki,J.M., Cochrane,G., Cole,J.R., Dinger,M.E., Enright,A.J., Gardner,P.P. *et al.* (2011) RNAcentral: a vision for an international database of RNA sequences. *RNA*, **17**, 1941–1946.